AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 110:95–104 (1999)

Estimating Sexual Dimorphism and Size Differences in the Fossil Record: A Test of Methods

J. A. REHG AND S.R. LEIGH*
Department of Anthropology, University of Illinois, Urbana, Illinois 61801

KEY WORDS fossils; primates; dimorphism estimates; dimorphism levels

ABSTRACT Investigations of size variation in fossil and archaeological skeletal assemblages may be complicated by incomplete skeletons, biased representation of sexes, and the lack of morphological features that identify sex. In order to refine our ability to evaluate size variation, we test the accuracy of three methods that are currently used to estimate size differences in unsexed (pooled) samples: the means method, the median method, and a newly applied technique, the method of moments. Using body mass data from 42 primate species, we calculated actual levels of sexual dimorphism for each species and compared these values to estimates produced by each method. Multivariate regression was used to examine the effects of sample distribution characteristics, including sample size, kurtosis, skewness, sample variance, sex ratio, and intrasexual variance on the performance of the methods. None of the methods appears to be especially accurate. However, one of the simplest methods, the means method, performs relatively well. Factors that lead to inaccuracies in estimation are not readily evident based on multiple regression analysis. We urge caution in the utilization of these techniques, and advocate further analysis of simulated data. Am J Phys Anthropol 110: 95–104, 1999.

Sex differences in size within primate species often reflect important and clearly defined social and ecological processes (Leutenegger and Cheverud, 1982; Leutenegger and Kelly, 1977; Plavcan and van Schaik, 1992, Clutton-Brock and Harvey, 1977). Correlations between dimorphism and these processes are crucial to the development of interpretations of behavior, ecology, and systematics in fossil forms (Borgognini Tarli and Repetto, 1986; Harvey et al., 1978; Kay et al., 1988; McHenry, 1986; Pickford, 1986a,b). Despite the importance of deriving accurate estimates of dimorphism in fossil assemblages, difficulties persist in applying and interpreting a variety of methods that serve to calibrate size differences in fossil assemblages. Although there are cases in which complete skeletons with reliable morphological indicators of sex are available for study [e.g., archaeological cases that enable use of pubic symphysis shape to estimate sex (Phenice, 1969)], the vast majority of situations involve incomplete remains and mixed skeletal samples. As a result, most estimates of fossil dimorphism must be based on the degree and patterning of size differences within assemblages of fossil materials (Kramer et al., 1995; Lieberman et al., 1988; Lockwood et al., 1996).

Several quantitative approaches to calibrating the level of size differences within fossil assemblages have been rigorously evaluated (e.g. Cope, 1993; Godfrey et al., 1993; Kramer and Konigsberg, 1999; Plav-

*Correspondence to: S. R. Leigh, Department of Anthropology, University of Illinois, 109 Davenport Hall, 607 S. Mathews Ave., Urbana, IL 61801. E-mail: s-leigh@uiuc.edu

Received 20 October 1998; accepted 22 May 1999.

can, 1994). The performance of these techniques varies; consequently, there is a premium on understanding what factors influence the reliability and accuracy of methods used to calibrate average size differences within mixed-sex distributions. In this paper, we evaluate several methods of inferring the level of size differences within fossil samples, and investigate the effects of "naturally occurring" variation in size distributions on these methods. Our analysis specifically focuses on an approach termed the method of moments technique (Josephson et al., 1996). This approach, initially developed in order to define subdistributions within mitochondrial DNA (mtDNA) mismatch distributions (see Rogers, 1997), has been proposed as an attractive alternative to previously developed methods (Josephson et al., 1996). According to Josephson et al. (1996), the method of moments is an unbiased estimator that is simple, allows construction of confidence intervals, and avoids reliance on living analogs. Evaluations of simulated data imply that the method of moments improved estimations of dimorphism when compared to some other techniques (Josephson et al., 1996).

In order to provide additional recommendations regarding approaches to estimating size differences within fossil assemblages, we compare the performance of the method of moments to previously proposed techniques (the means method and the median method). Our comparisons are restricted to these methods because they appear to be relatively accurate, provide estimates of dimorphism [rather than a referential measure of variation such as CV approaches (Cope, 1993; Plavcan, 1994)], and are simple to use. We compare these techniques using empirical data representing body masses of 42 primate species. In addition to comparing the accuracy of these techniques, we undertake multiple regression analyses of descriptors of pooled distributions in order to understand the determinants of deviations between actual and estimated levels of dimorphism for each technique. More specifically, multiple regression analyses, using descriptors of mixed-sex (pooled) distributions (skew, kurtosis, variance, and sample size), provide us with an empirical understanding of how the attributes of distributions influence estimators of size differences. Optimally, this approach would enable us to define characteristics of distributions that have major impacts on the accuracy of these estimators. Recommendations regarding the best choice of estimators can then be based on measurable aspects of distributions (e.g., skew, kurtosis, variance, and sample size).

Although recent comparisons of techniques have provided valuable insight into the capabilities of different methods to yield accurate estimates of size differences, we note that the data used in such studies are often simulated data. These data are typically sampled from normally distributed theoretical populations of infinite size, and may exhibit departures from normality (e.g., Plavcan, 1994; personal communication, 1999). However, it is difficult to define what sorts of departures from normality might be encountered in paleontological data. Thus, these theoretical samples may not be subject to the sources of bias seen in paleontological contexts. Therefore, we follow earlier approaches using actual data (Cope, 1993; Plavcan, 1994) for assessing the performance of the three estimation methods, and use multiple regression to examine the effects of variation in the properties of distributions on estimators of dimorphism. Analyzing the effects of "naturally occurring" departures from normality in these distributions is a necessary first step to refining simulation approaches to this problem.

MATERIALS AND METHODS

Materials

Body masses from 1,737 adult females and 1,134 adult males representing 42 anthropoid species, including New World monkeys, Old World monkeys, and apes, were used in this analysis (Table 1). These body mass data were collected by S.R.L. on live captive animals in normal, healthy condition (Leigh, 1992). Only data on adult individuals (defined as 1 year past age at growth cessation) were used (Leigh, 1992). Moreover, some individuals contributed more than one data point to a distribution so that

TABLE 1. A list of species used in the analysis of dimorphism estimation methods with actual and estimated values of sexual dimorphism according to each of the three methods tested

Species	Actual	MoM	Mean	Median
Callimico goeldii	0.988	1.257	1.322	1.304
Callithrix jacchus	0.939	1.132	1.229	1.224
Leontopithecus rosalia	1.079	1.324	1.191	1.200
Saguinus oedipus	0.938	1.247	1.153	1.151
Saguinus fuscicollis	1.001	1.201	1.226	1.226
Saguinus geoffroyi	1.001	1.179	1.167	1.157
Saguinus imperator	1.013	1.110	1.117	1.112
Aotus trivirgatus	0.927	1.682	1.238	1.236
Callicebus moloch	1.130	1.506	1.353	1.342
Cebus apella	2.100	1.879	1.856	1.839
Saimiri sciureus	1.394	1.311	1.428	1.410
Alouatta caraya	1.522	2.202	2.190	2.190
Ateles geoffroyi	1.342	1.299	1.300	1.299
Allenopithecus nigroviridis	1.733	2.299	1.050	1.050
Miopithecus talapoin	1.335	1.466	1.462	1.439
Erythrocebus patas	1.920	1.931	1.953	1.819
Cercopithecus aethiops	1.670	2.289	1.718	1.718
Cercopithecus diana	1.833	1.860	1.833	1.356
Cercopithecus mitis	1.802	1.470	1.768	1.666
Cercopithecus neglectus	1.949	1.829	1.949	1.655
Macaca arctoides	1.530	1.362	1.518	1.452
Macaca cyclopis	1.358	1.439	1.510	1.515
Macaca fascicularis	1.593	1.929	1.577	1.278
Macaca fuscata	1.443	1.373	1.452	1.452
Macaca mulatta	1.691	1.219	1.516	1.184
Macaca nemestrina	2.053	2.086	2.053	2.002
Macaca radiata	1.822	1.876	1.801	1.739
Macaca silenus	1.686	1.732	1.757	1.757
Mandrillus sphinx	2.116	2.15	1.911	1.740
Mandrillus leucophaeus	2.251	2.536	2.251	2.088
Cercocebus torquatus	1.600	1.602	1.440	1.345
Papio hamadryas	1.647	1.798	1.647	1.601
Colobus quereza	1.360	1.392	1.397	1.400
Presbytis (Trachypithecus) cristata	1.103	1.188	1.280	1.265
Presbytis (Trachypithecus) obscura	1.096	1.239	1.133	1.131
Presbytis (Trachypithecus) entellus	1.484	1.408	1.499	1.480
Pygathrix nemaeus	1.490	1.490	1.512	1.508
Hylobates lar	1.052	1.285	1.280	1.278
Hylobates syndactylus	1.056	1.212	1.280	1.137
Gorilla gorilla	1.960	1.897	1.956	1.872
Pan paniscus	1.325	1.000	1.335	1.331
Pan troglodytes	1.265	1.420	1.307	1.331

sample sizes range from 9 to 252. While this may affect statistical independence, analyzing the data in this way increases the range of variation in sample sizes, outweighing costs of some loss of independence. Actual body size dimorphism levels in these species were calculated by a ratio measure (average male mass/average female mass) (Table 1). Body mass data were then disassociated from sex and used to estimate dimorphism in all 42 species using the method of moments, median method, and mean method.

Estimators of dimorphism

Method of moments. Josephson et al. (1996) describe the exact process for estimating dimorphism by the method of moments. This technique is a form of finite mixture analysis (Dong, 1997; Godfrey et al., 1993; Kramer and Konigsberg, 1999). Briefly, the method of moments technique estimates the theoretical underlying male and female distributions from a combined male-female (pooled) sample based on observations in that sample. The theoretical moments are estimated from these data, and the level of dimorphism is calculated based on the means of the theoretical male-female distributions. A related technique is applied to evaluating moments in DNA mismatch distributions. Josephson et al.'s method is extremely straightforward. Technically, the method involves calculating z-scores based on logged values for each individual in a sample. These values are raised to the fourth power, and the average of these values estimated. This average (m_4) is substituted into the equation:

$$d^{(z)} = ((3/2) - (m_4/2))^{1/4} \tag{1}$$

The value $d^{(z)}$ is then multiplied by the ln-transformed standard deviation ("undoing" the z-transformation) to yield a value termed d. Estimating sexual dimorphism is calculated by the formula

$$e^{2(d)} \tag{2}$$

Median method. The median method involves determining the median of a pooled or unknown distribution, then splitting the sample at this value into a set of values greater than the median and a set lesser than the median (as described by Plavcan, 1994; Josephson et al., 1996). For primates, and most mammals, the "greater-than" set presumably represents males, and the "lesser-than" set represents females. The means for each of these sets are calculated, then the "female" mean is divided into the "male" mean, resulting in an estimate of sexual dimorphism. Previous research suggests that this approach is less reliable than the means method (Plavcan, 1994).

Means method. The means method is identical to the median method, except that the undifferentiated sample is divided into two sets of values at the mean of the total sample, rather than the median (Plavcan, 1994). The means of these divided sets are then used to calculate the dimorphism estimate, as with the median method. Plavcan finds that this method is comparatively reliable, but like other methods, is influenced by sex ratio and intrasexual variation.

Comparison procedures

The differences between actual and predicted values of sexual dimorphism using all three methods were calculated (actual-estimated), as were the absolute differences in estimates and actual values (|actual-estimated|). The first phase of the analysis uses nonparametric Lowess regressions (Efron and Tibshirani, 1991) for visual representations of the general pattern by which actual and estimated measures of dimorphism differ. Lowess regressions are not bound by model form (e.g., linear, quadratic) but should illustrate patterned departures of estimated from actual values. In addition, we plot a vector of slope = 1, intercept of 0, to illustrate the deviation of estimated points from expected actual values. We subsequently utilize multiple regression analysis to evaluate the influence of variation in the properties of distributions on the accuracy of estimated dimorphism. Specifically, we examine the effects of independent variables that describe the actual distributions on which the estimates of dimorphism are based. These variables include sample size, kurtosis, skewness, and variance. Kurtosis and skewness are measures of departures from normality. Skewness, the third moment of a distribution, reflects the degree to which a tail of a distribution is "drawn out" (Sokal and Rolhf, 1981:114). Kurtosis, the fourth moment of a distribution, measures the "peakedness" of a curve (Sokal and Rolhf, 1981:114). We then perform multivariate least-squares regressions using these moments as predictors of the difference between actual and estimated dimorphism for all three methods, as well as the absolute value of the difference between actual and estimated dimorphism for all three methods.

The third and fourth moments (skew and kurtosis, respectively), along with sample size and variance, were chosen as independent variables because they describe characteristics that can be measured directly from a fossil sample. These moments also provide information on whether or not the sample approximates a normal distribution. Josephson et al. (1996), in their initial computer simulation analysis of the method of moments, note that small sample sizes (below 50 specimens), a biased sex ratio, and high intrasexual variance negatively affect the accuracy of this method. Similarly, Plavcan (1994) notes that small sample size, biased sex ratio, and high intrasexual variance impaired the accuracy of the four methods he tested. Because the composition of fossil samples is unknown, we prefer to use simple descriptors of distributions rather than estimates of sex ratio and intrasexual variance, which cannot be known in fossil studies. For example, a highly biased sex ratio will theoretically skew a distribution in one direction. While the sex ratio cannot be ascertained from a fossil sample, the degree of skewness can be calculated directly. Our actual data are subject to the same kinds of problems routinely found in skeletal samples; all of our independent variables vary to some degree, while simulated studies may not have incorporated variation in moments among samples. However, in order to examine the possible effects of intrasexual variance and sex ratio on these methods, and facilitate comparison of this study with previous tests of estimators, we also include these two variables in a second set of multivariate regression analyses. The sex ratio was calculated for each species by dividing the sample size of the sex with more individuals by the sample size of the sex with fewer individuals. This provided an absolute measure of the difference in sample sizes between males and females. A ratio of intrasexual variance was calculated for each species by dividing the variance in the sample of males by variance in the sample of females of a species.

Significant relationships between one or more of these sample or distribution charac-

teristics (independent variable) and the difference between actual and estimated levels (dependent variable) would indicate that the independent variable accounts for variation in a particular method's accuracy. Moreover, the direction in which a particular method erred (either underestimated or overestimated) can be inferred. Significant relationships between an independent variable and the absolute difference between actual and estimated dimorphism would indicate that the variable can explain variation in a particular method's accuracy overall, although not whether the estimator under- or overestimated. Interactions among independent variables may influence the effect of a variable on method accuracy. In order to investigate potential interactions, the accuracy of each method was regressed on all two-variable interactions (i.e., sample size by variance; sample size by kurtosis; sample size by skewness, etc.). Given that we have no a priori hypotheses as to how these variables should interact, these analyses are considered exploratory. We adopt $p = .05$ as a measure of statistical significance. All analyses were conducted in Excel and Systat.

RESULTS

Qualitative analysis

All three methods tended to overestimate sexual dimorphism below approximately 1.5 units of actual dimorphism (Fig. 1). In contrast, the three methods tended to underestimate sexual dimorphism at high actual dimorphism levels (above 1.5 units). The median method seems to be slightly more prone to overestimation and underestimation in these ranges than other estimators. In contrast, the means method was the most consistent in overestimating dimorphism levels, although it underestimated actual levels in 11 of 42 cases (Table 2). Overall, the means method produced the closest estimates to actual levels, but like other methods, the means method produced estimates that were occasionally highly divergent. At medium to high dimorphism levels in particular, estimates with any of the three methods were inaccurate up to 25% or more, with the method of moments exhibiting the most divergent and variable estimates of all three methods. On the other hand, the re-

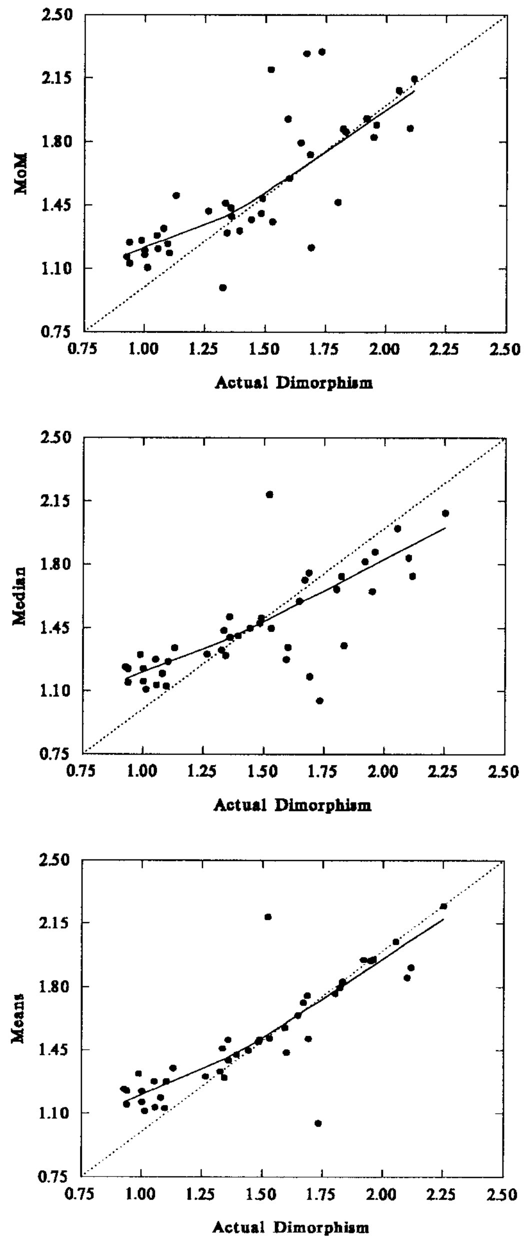

Fig. 1. Estimates of sexual dimorphism from the method of moments, means method, and median method and their relationship to actual levels of dimorphism. All methods systematically overestimate dimorphism at the lower end of the range. The means method appears to provide the best overall fit to actual estimates. Extreme outliers include *Allenopithecus nigroviridis* (lowest data point in the means plot) and *Alouatta caraya* (high data point in the middle range of the means plot). Both estimates are based on comparatively small sample sizes.

TABLE 2. Estimates of overall method performance[1]

Performance	Method of moments	Means method	Median method
Percent average departure from actual	20%	13%	18%
Number of overestimates (%)	30 (72%)	26 (62%)	24 (18%)
Number of estimates = actual (%)	1 (2%)	5 (12%)	0 (0%)
Number of underestimates (%)	11 (26%)	11 (26%)	18 (43%)
Totals	42 (100%)	42 (100%)	42 (100%)

[1] Performance measures include percent departure from actual dimorphism and numbers of times that methods overestimate dimorphism, estimate it exactly, and underestimate dimorphism.

gression of the method of moments estimates suggest that this method most accurately estimates dimorphism in the entire assemblage of species. Worst performance cases tended to occur for samples characterized by actual dimorphism levels in the middle ranges. Moreover, variability in each method's accuracy tended to be larger for samples with higher actual dimorphism levels. Averaging estimates from all three methods failed to increase prediction reliability.

Summarizing accuracy of the various methods suggests that the means method tends to be more accurate than other methods (Table 2). The means method shows the lowest average absolute departure from actual dimorphism. Moreover, in 12% of the cases, the means method returned a value identical to the actual estimate of dimorphism. All methods have a tendency to overestimate dimorphism in the low ranges of dimorphism, as indicated in Lowess regression plots (Fig. 1).

Regression analyses

Characteristics of distributions. Independent variables that describe distributions are minimally intercorrelated (Table 3). Multivariate regressions with four independent variables (sample size, kurtosis, skew, and variance) suggest that variation in the method of moments was only slightly influenced by these variables (Table 4), but neither the accuracy of the means nor median method was affected by these measures. In these analyses, directional signs (positive or negative) were not retained for skew and kurtosis. Variation in the accuracy of the method of moments was significantly affected only by skew, which implies that the method is influenced when a distribution possesses a long tail (Table 4). More specifically, given relatively unskewed distributions, method of moments tended to overestimate dimorphism, but given very skewed distributions (either positively or negatively), method of moments tended to underestimate dimorphism. It should be noted that the overall regression is not significant as a predictive model of the method's accuracy. Furthermore, as is expected with such weak regression results, individual samples did not necessarily conform to the trends. For example, dimorphism levels for very skewed individual cases [e.g., *Callimico goeldii*, *Presbytis* (*Trachypithecus) cristata*] were overestimated using the method of moments, despite a general trend to underestimate dimorphism in skewed samples. The absolute degrees of error in estimates by the method of moments, means method, or median method were not affected by any of the sample distribution variables (Table 4). We could not identify a consistent effect of the sample attributes on the performance of the methods, nor can we determine the "best" estimation method given certain sample and distribution characteristics.

Intrasexual variance and sex ratio. Including information about intrasexual variance and sex ratio substantially increased the amount of variation explained by multivariate regression models (Table 5). Variation in the accuracy of the method of moments and the means and median methods was unaffected by sex ratio. Similarly, variation in the absolute deviation of each method from actual values was not significantly affected by sex ratio. In contrast, the intrasexual variance ratio had a significant effect on the accuracy of the means method, but did not affect the median method or method of moments. However, variation in the absolute deviation of all three methods from actual values was significantly affected by relative amounts of intrasexual variation in males and females (Table 5). Specifically, as the variance in the male sample increased relative to the variance in the female sample, dimorphism estimates of the

TABLE 3. Pearson product-moment correlation measuring the association among moments of distributions[1]

	Sex ratio	Kurtosis	Skewness	Standard deviation	Variance	Range
Sex ratio	1.00					
Kurtosis	−0.37	1.00				
Skewness	−0.47	0.34	1.00			
Standard deviation	−0.07	−0.23	0.10	1.00		
Variance	−0.09	−0.17	0.12	1.00	1.00	
Range	−0.04	−0.15	0.05	0.95	0.94	1.00

[1] Correlations between sex ratio and other measures are also included.

TABLE 4. Multivariate regression results[1] *evaluating the effects of four independent variables (sample size, variance, absolute kurtosis, and absolute skewness) on accuracy of the method of moments, the means method, and median method*

Test	Method	R^2	Variable(s) with significant effect(s)
Accuracy	MoM[2]	0.244[3]	Absolute skewness
Accuracy	Means	0.032	None
Accuracy	Median	0.097	None
Absolute deviation	MoM	0.000	None
Absolute deviation	Means	0.000	None
Absolute deviation	Median	0.000	None

[1] Only coefficients that have significant effects ($p < 0.05$) on accuracy (deviation between actual and predicted) and absolute deviation (absolute value of deviation) are reported.
[2] MoM represents the method of moments.
[3] Significant regression ($p < 0.05$).

TABLE 5. Multiple regression results[1] *evaluating the effects of six independent variables (sample size, variance, absolute kurtosis, absolute skewness, sex ratio, and intrasexual variance ratio) on accuracy of the method of moments, the means method, and median method*

Test	Method	R^2	Variable(s) with significant effect(s)
Accuracy	MoM[2]	0.288[3]	Intrasexual variance ratio
Accuracy	Means	0.343[3]	None
Accuracy	Median	0.270[3]	None
Absolute deviation	MoM	0.165	Intrasexual variance ratio
Absolute deviation	Means	0.372[3]	Intrasexual variance ratio
Absolute deviation	Median	0.213[3]	Intrasexual variance ratio

[1] Only coefficients that have significant effects ($p < 0.05$) on accuracy (deviation between actual and predicted) and absolute deviation (absolute value of deviation) are reported.
[2] MoM represents the method of moments.
[3] Significant regression ($p < 0.05$).

method of moments, means method, and median method deviate more from actual values. Despite this significant trend, estimates for individual cases characterized by high ratios of male to female intrasexual variance did not conform to this general relationship.

Interactions. None of the several two-variable interactions that we investigated had a significant effect on variability in method accuracy. Given that we constructed regression models with either four or six independent variables, we must note that large numbers of interactions are possible. In the absence of a clear theory of how interactions should affect these estimators, we have analyzed only a small number of possible interactions, preferring to refrain from unguided exploration of the data.

DISCUSSION

Studying sexual dimorphism in fossil primate taxa is important for interpreting taxonomy, anatomy, and also potentially the ecology and social behavior of extinct primates (Plavcan and van Shaik, 1992; Richmond and Jungers, 1995). In terms of taxonomy, accurate and precise estimates of sexual dimorphism of fossil taxa provide a way for researchers to evaluate variability within a fossil assemblage, and determine if those levels are more consistent with intra- or interspecific levels of variation (Cameron, 1991, 1992; Josephson et al., 1996; Teaford et al., 1993). Despite the potential value of the methods we evaluated, qualitative and quantitative analyses suggest that no one method provides especially congruent estimates of dimorphism (Godfrey et al., 1993; Plavcan, 1994). For instance, our qualitative analyses, such as inspection of scatterplots, clearly show that arriving at a reliable estimate for a single case appears problematic, especially in the middle to upper ranges of dimorphism for all three methods. On the other hand, moderate levels of dimorphism appear to be most accurately measured, but

cases of low dimorphism are consistently overestimated. Downward adjustment of low estimates could improve the accuracy of estimation, but this obviously can only be done on an ad hoc basis.

Comparisons by Josephson et al. (1996) indicate the method of moments provides conservative estimates of dimorphism, improving on estimation capabilities of other methods. We find that the method of moments does not provide numerous advantages to estimation of dimorphism with our data over either the means or median method. We emphasize that this finding has important implications for analyses that attempt to define subdistributions within DNA mismatch and intermatch distributions with the method of moments [see also Rogers' cautious assessment of this method in this context (1997)]. In addition, the means method seems to be at least marginally more accurate and consistent than the other methods. Clearly, further research into this problem seems warranted.

Regression analyses that examine observable attributes of distributions do little to clarify why estimates provided by various techniques depart from actual values. Specifically, variability in the accuracy of the means and median methods is unaffected by the variables analyzed. As noted, skewness has a significant effect on the accuracy of the method of moments, but this effect is quite weak and may be trivial. Consequently, none of these sample characteristics could be used to predict either the approximate accuracy of a particular method or the direction in which a method's prediction departs from the actual value.

Unfortunately, variability in the accuracy of the means and median methods was not affected by any of these features, and only skewness had a significant effect on the accuracy of the method of moments. Additionally, although we found a significant negative relationship between method of moments estimates and skewness, there were still sample cases characterized by extreme skew for which the method of moments overestimated dimorphism. Skewness was thus not a reliable indicator of how the method of moments performed for individual samples. As a result, none of these sample characteristics could be used to predict either the approximate accuracy of a particular method, or the tendency to under- or overestimate dimorphism for specific samples.

Our results are consistent with previous studies finding that intrasexual variance distorts estimates of fossil dimorphism (Godfrey et al., 1993; Josephson et al., 1996; Plavcan, 1994). Biased sample composition, as represented by skewness or the intrasexual variance ratio, influenced the accuracy of all methods (although this depended upon how accuracy was measured, i.e., absolute versus direction of error). However, the effects of intrasexual variance bias are subtle. It seems to influence accuracy even though it is not directly reflected in the shapes of distributions, at least in terms of skew and kurtosis. The observation that kurtosis and skewness had no significant effect on the accuracy of these two methods despite the significant influence of intrasexual variance demonstrates that kurtosis and skewness do not perfectly reflect biases in sex ratio and intrasexual variance. In fact, bias in representation of males and females among fossils is a distinct possibility, and may thus lead to errors in estimates of size differences within an assemblage if each sex is assumed to be equally represented and characterized by similar levels of intrasexual variance.

Although Josephson et al. (1996) demonstrated that sample size, intrasexual variance, and sex ratio negatively affect the accuracy of the method of moments, we did not find a significant relationship between variation in sample size, sex ratio, and variation in method accuracy. Methodological differences may be partly responsible for this discrepancy. Like Plavcan (1994), Josephson et al. (1996) used computer simulations to evaluate the performance of the method of moments, and were thus able to isolate the effects of sample size while holding other variables constant. Our approach did not allow this strict control, and interactions between sample size and other features may have mediated or obscured the effect of sample size on the accuracy of the method of moments. Although we did not find any significant interactions that influence perfor-

mance of the three estimators, we only examined a limited number of two-variable interactions. It is possible that more complex multivariate interactions disrupt method performance. Alternatively, our sample size was not large enough to detect such interactions. We should emphasize that the lack of a theoretical model explaining interactions limits further research in this area.

CONCLUSIONS

Direct comparison of these three methods for estimating dimorphism suggests that all are roughly equal in terms of accuracy. We are hesitant to enthusiastically recommend any of these procedures for estimating fossil dimorphism. However, the means method seems to provide an extremely simple and convenient way to estimate dimorphism from assemblages of unsexed fossil materials. At low levels of dimorphism, the means method appears to inflate estimates of dimorphism. At higher levels of dimorphism, large differences in sample sizes on either side of the mean may affect the accuracy of the method, and a proper dose of skepticism can be applied to the resulting estimate. Further analyses with simulated data are crucial to further exploration of this problem. Importantly, such analyses should focus on interactions among variables and should relax the assumption of normal distributions. Obviously, when using any of these estimation methods to determine size differences on unsexed specimens, the results should be interpreted with caution.

ACKNOWLEDGMENTS

An abbreviated version of this paper was presented at the 1997 meetings of the American Associations of Physical Anthropologists. We thank Steven Josephson for helpful comments on that presentation. Drs. Mike Plavcan and William Jungers provided helpful comments on an earlier version of this paper.

LITERATURE CITED

Bohrnstedt GW, Knoke D. 1982. Statistics for social data analysis. Itasca, IL: FE Peacock Publishers.

Borgognini Tarli SM, Repetto E. 1986. Methodological considerations on the study of sexual dimorphism in past human populations. In: Pickford M, Chiarelli B, editors. Sexual dimorphism in living and fossil primates, Giugno: Il Sedicesimo. p 51–66.

Cameron DW.1991. Sexual dimorphism in the early Miocene species of *Proconsul* from the Kisingiri formation of East Africa: A morphometric examination using multivariate statistics. Primates 32:329–343.

Cameron DW. 1992. A morphometric analysis of extant and early Miocene fossil hominoid maxillo-dental specimens. Primates 33:377–390.

Clutton-Brock TH, Harvey PH. 1977. Primate ecology and social organization. J Zool (Lond) 183:1–39.

Cope DA. 1993. Measures of dental variation as indicators of multiple taxa in samples of sympatric *Cercopithecus* species. In: Kimbel WH, Martin LB, editors. Species, species concepts, and primate evolution. New York: Plenum Press. p 211–238.

Dong Z. 1997. Mixture analysis and its preliminary application in archaeology. J Archeol Sci 24:141–161.

Efron B, Tibshirani R.1991. Statistical data analysis in the computer age. Science 253:390–395.

Godfrey LR, Lyon SK, Sutherland MR. 1993. Sexual dimorphism in large-bodied primates: the case of the subfossil lemurs. Am J Phys Anthropol 90:315–334.

Harvey PH, Kavanagh M, Clutton-Brock TH. 1978. Sexual dimorphism in primate teeth. J Zool (Lond) 186:475–485.

Josephson SC, Juell KE, Rogers AR. 1996. Estimating sexual dimorphism by method of moments. Am J Phys Anthropol 100:191–206.

Kay RF, Plavcan JM, Glander KC, Wright PC. 1988. Sexual selection and canine dimorphism in New World monkeys. Am J Phys Anthropol 77:385–397.

Kelley J. 1995. Sex determination in Miocene Catarrhine primates. Am J Phys Anthropol 96:391–417.

Kramer A, Donnelly SM, Kidder JH, Ousley SD, Olah SM. 1995. Craniometric variation in large-bodied hominoids: Testing the single species hypothesis for *Homo habilis*. J Hum Evol 29:443–462.

Kramer A, Konigsberg LW. 1999. Recognizing species diversity among large-bodied hominoids: a simulation test using missing data finite mixture analysis. J Hum Evol 36:409–421.

Leigh SR. 1992. Patterns of variation in the ontogeny of primate body size dimorphism. J Hum Evol 23:27–50.

Leutenegger W, Cheverud J. 1982. Correlates of sexual dimorphism in primates: ecological and size variables. Int J Primatol 3:387–402

Leutenegger W, Kelly JT. 1977. Relationship of sexual dimorphism in canine size and body size to social, behavioral, and ecological correlates in anthropoid primates. Primates 18:117–136.

Lieberman DE, Pilbeam DR, Wood BA. 1988. A probabilistic approach to the problem of sexual dimorphism in *Homo habilis*: a comparison of KNM-ER 1470 and KNM-ER 1813. J Hum Evol 17:503–511.

Lockwood CA, Richmond BG, Jungers WL, Kimbel WH. 1996. Randomization procedures and sexual dimorphism in *Australopithecus afarensis*. J Hum Evol 31:537–548.

McHenry HM. 1986. Size variation in the postcranium of *Australopithecus afarensis* and extant species of Hominoidea. In: Pickford M, Chiarelli B, editors. Sexual dimorphism in living and fossil primates. Giugno: Il Sedicesimo. p 183–189.

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. Am. J Phys Anthropol 30:297–302.

Plavcan JM. 1994. Comparison of four simple methods for estimating sexual dimorphism in fossils. Am. J Phys Anthropol 94:465–476.

Plavcan JM, van Shaik CP. 1992. Intrasexual competition and canine dimorphism in anthropoid primates. Am J Phys Anthropol 87:461–477.

Pickford M. 1986a. Sexual dimorphism in *Proconsul*. In:

Pickford M, Chiarelli B, editors. Sexual dimorphism in living and fossil primates. Giugno: Il Sedicesimo. p 133–170.

Pickford M. 1986b. On the origins of body size dimorphism in primates. In; Pickford M and Chiarelli B, editors. Sexual dimorphism in living and fossil primates. Giugno: Il Sedicesimo. p 77–91.

Richmond BG, Jungers WL. 1995. Size variation and sexual dimorphism in *Australopithecus afarensis* and living hominoids. J Hum Evol 29:229–245.

Rogers AR. 1997. Population structure and modern human origins. In: Donnelly P, Tavaré S, editors. Progress in population genetics and human evolution. New York: Springer-Verlag. p 55–79.

Sokal RR, Rohlf FJ. 1969. Biometry. San Francisco: W H Freeman.

Teaford MF, Walker A, Mugaisi GS. 1993. Species discrimination in *Proconsul* from Rusinga and Mfangano Islands, Kenya. In: Kimbel WH, Martin LB, editors. New York: Plenum Press, p 393–427.